On estimating stock market volatility and crash risk

David S. Bates, Finance

ACMS Lecture

May 4, 2012

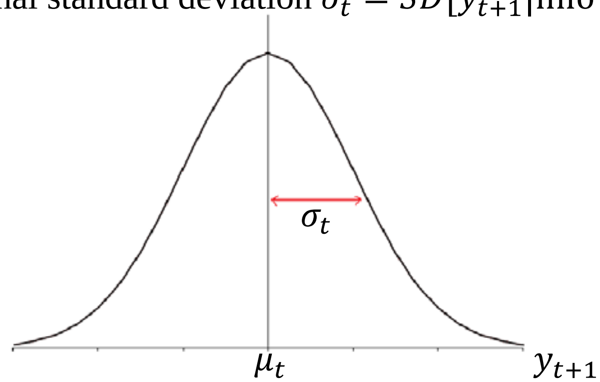
Outline

- Issue: estimating conditional distributions of stock market returns
- Early methodologies
- Filtration
- Challenges
- Approaches

Conditional distributions of financial returns y_{t+1}

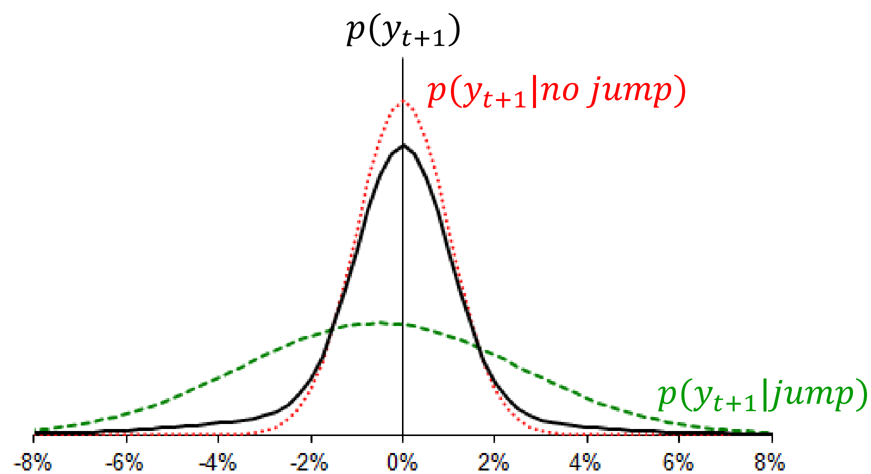
If Gaussian, distribution summarized by

- conditional mean $\mu_t = E[y_{t+1}|\text{info at time } t] \approx 0$ daily
- conditional standard deviation $\sigma_t = SD[y_{t+1}|\text{info at time } t]$



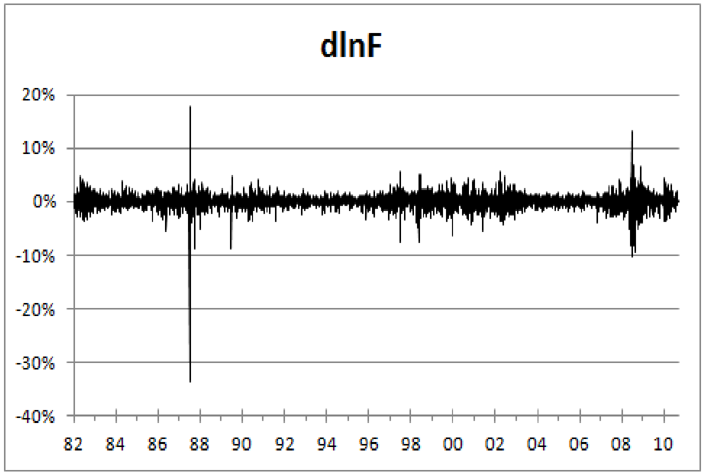
Conditional distributions of financial returns y_{t+1}

Non-Gaussian distributions are often modeled as a mixture of Gaussians.



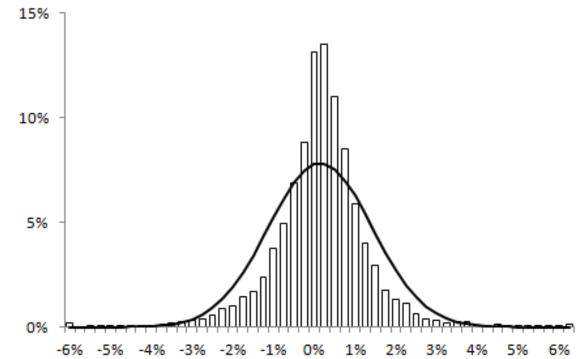
Data: daily log-differenced S&P 500 futures prices

(roughly S&P percentage returns in excess of riskless T-bill rate)



Important properties of stock market returns

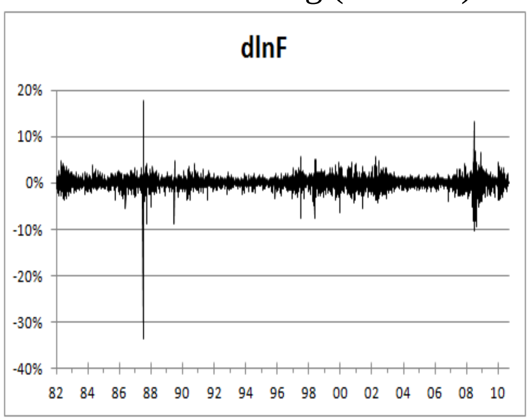
1) Unconditional and conditional distributions are not Gaussian. There are major outliers.



Mean: 0.02%
SD: 1.27%
Max: 17.7%
Min: -33.7%
Skew: -2.36
Xkurt: 80.9

(log futures returns over 1982-2010)

2) We observe *clusters* of big (or small) absolute returns.



3) Markets are more volatile following market drops than following comparable market increases.

Some early moving-average approaches to modeling persistence in conditional variances of stock market returns

- Moving-window sample variance
- Riskmetrics (J.P. Morgan)
- GARCH

Notation

$y_{t+1} = \ln(F_{t+1}/F_t)$ is the log-differenced futures price

$\sigma_t = SD(y_{t+1} | \text{information available at time } t)$

Moving-window sample variance

$$\hat{\sigma}_t^2 \approx \frac{1}{N}\sum_{n=1}^{N} y_{t+1-n}^2 \quad \text{(ignoring } \mu_t\text{, which is} \approx 0 \text{ daily)}$$

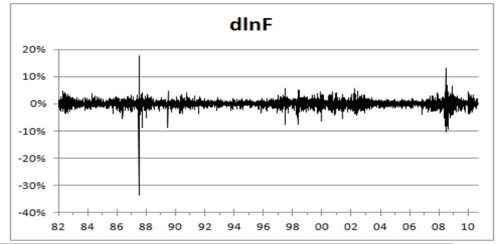
RiskMetrics

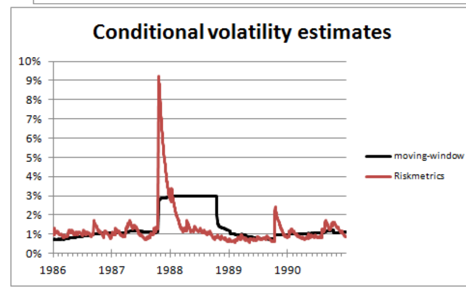
$\hat{\sigma}_t^2 \approx \sum_{n=1}^{\infty} w_n y_{t+1-n}^2$ for $w_n = .06(0.94)^n$

Generalized autoregressive conditional heteroskedasticity (GARCH(1,1)):

$$\hat{\sigma}_t^2 - \theta \approx \alpha(y_t^2 - \theta) + \beta(\sigma_{t-1}^2 - \theta), \text{ where } \theta = E[\sigma_t^2].$$

$\hat{\sigma}_t$: 252-day sample volatility versus RiskMetrics





Moving-window $\hat{\sigma}_t$ drops when '87 outliers leave the window, in 1988.

All of these are ***filtration algorithms*** $\hat{\sigma}_t = f(\hat{\sigma}_{t-1}, y_t, \ldots)$ that capture substantial persistence in $\hat{\sigma}_t$.

- Moving-window: $\hat{\sigma}_t^2 \approx \hat{\sigma}_{t-1}^2 + \frac{1}{N}(y_t^2 - y_{t-N}^2)$
- RiskMetrics: $\hat{\sigma}_t^2 \approx 0.94\, \hat{\sigma}_{t-1}^2 + 0.06 y_t^2$
- GARCH(1,1): $\hat{\sigma}_t^2 - \theta \approx \alpha(y_t^2 - \theta) + \beta(\sigma_{t-1}^2 - \theta)$
  (includes RiskMetrics for $\theta = 0$, $\alpha + \beta = 1$)

Enormous variety of GARCH specifications

$\hat{\sigma}_t = f(\hat{\sigma}_{t-1}, y_t, \ldots)$

-additional lags

-different functional forms of f

-can allow different $\hat{\sigma}_t$ updating for $y_t > 0$ versus <0.

ML estimation of parameters under the assumption

$y_{t+1} = \mu + \sigma_t \tilde{\varepsilon}_{t+1}$

for $\tilde{\varepsilon}_{t+1}$ i.i.d. from some distribution, possibly fat-tailed.

σ_t is assumed *known* given the parameters & past data.

Stochastic volatility models of the joint evolution of (y_t, σ_t)

Discrete-time

$$y_{t+1} = \mu + \sigma_t \varepsilon_{t+1}$$
$$\ln \sigma_{t+1}^2 = \omega + \phi \ln \sigma_t^2 + \eta_{t+1}$$

Continuous-time

$$d \ln F_t = (\mu_0 + \mu_1 \sigma_t^2)dt + \sigma_t dW_{1t} + \gamma dN_t$$
$$d\sigma_t^2 = (\alpha - \beta \sigma_t^2)dt + \sigma_t dW_{2t}$$

where dW_{1t}, dW_{2t} are Wiener shocks with correlation $\rho < 0$;

dN_t is a Poisson counter with intensity $\lambda_t = \lambda_0 + \lambda_1 \sigma_t^2$ that counts the occurrence of jumps of size $\gamma \sim N(\bar{\gamma}, \delta^2)$; $Prob_t(dN_t = 1) = \lambda_t dt$.

(1-factor model of volatility *and* tail risk)

Filtration

$p(y_{t+1}, \sigma_{t+1} | y_t, \sigma_t)$: joint transition density from time series model conditional upon knowing σ_t (which we don't)

$\boldsymbol{Y}_t = \{y_1, y_2, \ldots, y_t\}$: set of observed data
$p(\sigma_t | \boldsymbol{Y}_t)$: what is known about σ_t at time t (including $E[\sigma_t | \boldsymbol{Y}_t]$)

Then

$p(y_{t+1}, \sigma_{t+1} | \boldsymbol{Y}_t)$: joint density conditional on $\boldsymbol{Y}_t$
$p(y_{t+1} | \boldsymbol{Y}_t)$: marginal density of y_{t+1}conditional on $\boldsymbol{Y}_t$

can be evaluated via

$$p(y_{t+1}, \sigma_{t+1} | \boldsymbol{Y}_t) = \int p(y_{t+1}, \sigma_{t+1} | y_t, \sigma_t) p(\sigma_t | \boldsymbol{Y}_t) d\sigma_t$$

Filtration uses Bayes' rule to update $p(\sigma_t | \boldsymbol{Y}_t)$ recursively over time, given the latest observation y_{t+1}:

$$p(\sigma_{t+1} | \boldsymbol{Y}_{t+1}) = \frac{p(y_{t+1}, \sigma_{t+1} | \boldsymbol{Y}_t)}{p(y_{t+1} | \boldsymbol{Y}_t)}$$
$$= \frac{\int p(y_{t+1}, \sigma_{t+1} | y_t, \sigma_t) p(\sigma_t | \boldsymbol{Y}_t) d\sigma_t}{\int p(y_{t+1} | y_t, \sigma_t) p(\sigma_t | \boldsymbol{Y}_t) d\sigma_t}$$

I.e., $p(\sigma_t | \boldsymbol{Y}_t)$ & $y_{t+1} \rightarrow p(\sigma_{t+1} | \boldsymbol{Y}_{t+1})$

Challenges/difficulties

- How do we ***summarize*** the entire univariate function $p(\sigma_t | \boldsymbol{Y}_t)$ across all values of σ_t at each point in time *t*?
- How do we update an entire function, and summarize the result?

1) Approaches: special analytical cases

Kalman filtration

1. If observed y_t and latent x_t have a jointly Gaussian evolution, $x_t | \boldsymbol{Y}_t$ is Gaussian and is summarized by its first two moments $E(x_t | \boldsymbol{Y}_t), Var(x_t | \boldsymbol{Y}_t)$.
2. Updating the moments is regression-like and simple.

Markov chains

1. x_t can take on a finite number of values $\{x^1, \ldots, x^S\}$, with filtered probabilities $\pi_{t|t}^s = Prob[x_t = x^s | \boldsymbol{Y}_t]$, $s = 1, \ldots, S$.
2. Updating these probabilities involves summation rather than integration.

2) Approximate approaches

Use these approaches as approximations

- "unscented", "robust" Kalman filtrations: summarize what is known about x_t as Gaussian even when structure is not Gaussian.
  -update its moments over time.
- discrete-state approximations to a continuously-distributed latent variable x_t.

3) Monte Carlo filtration (particle filter)

1) Summarize $p(x_t|\boldsymbol{Y}_t)$ by a large number (e.g., 10,000) of draws from that density -- a random histogram.

2) Update to $p(x_{t+1}|\boldsymbol{Y}_{t+1})$ by a large number of draws from $p(x_{t+1}|x_t)$, with heavier posterior probability weights placed on those draws that make observed y_{t+1} more likely (relatively high $p(y_{t+1}|x_{t+1})$).

3) Go to step 1.

4) Various refinements (resampling, auxiliary particle filters) to deal with distributions with low-probability outliers.

4) My approach: Bayes' rule for characteristic functions

Define $F(i\Phi, i\psi) = \mathrm{E}\left[\mathrm{e}^{i\Phi y + i\psi x}\right] = \iint e^{i\Phi y + i\psi x}\, p(y,x)\,dy\,dx$ as the joint CF (and Fourier transform) of *y* and *x*.

The characteristic function $G(i\psi|y) = E[e^{i\psi x}|y]$ of *x* *conditional* upon *y* can be computed via inverse Fourier transforms of *F*:

$$G(i\psi|y) = \frac{\int F(i\Phi, i\psi)e^{-i\Phi y}d\Phi}{\int F(i\Phi, 0)e^{-i\Phi y}d\Phi}$$

Posterior moments: $E[x^m|y] = \left.\frac{\partial^m G(\psi|y)}{\partial \psi^m}\right|_{\psi=0}$

Applied using a form of robust Kalman filtration

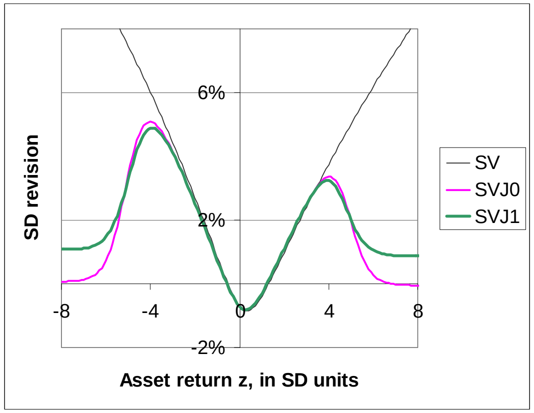
- Start with $G_{t|t}(i\psi) = E[e^{i\psi x_t}|\boldsymbol{Y}_t]$ associated w/ $p(x_t|\boldsymbol{Y}_t)$
- Update posterior mean & variance of x_{t+1} given y_{t+1}
- Use two-moment approximate distribution for $G_{t+1|t+1}(i\psi)$
- Continue

Advantages

- Don't typically have analytical $p(y_{t+1}, x_{t+1}|y_t, x_t)$ or $p(y_{t+1}, x_{t+1}|\boldsymbol{Y}_t)$ But for affine or quadratic time series models, we ***do*** have their joint characteristic functions.
- Can be used both for estimating time series parameters and for filtering latent variables.

Model-specific implications for volatility filtration

If y_{t+1} jumps and σ_{t+1} doesn't, it is optimal to ***downweight*** large $|y_{t+1}|$'s impact on $\hat{\sigma}_{t+1} - \hat{\sigma}_t$ revisions. (Bates, *RFS* 2006)



SV: no y_{t+1} jumps

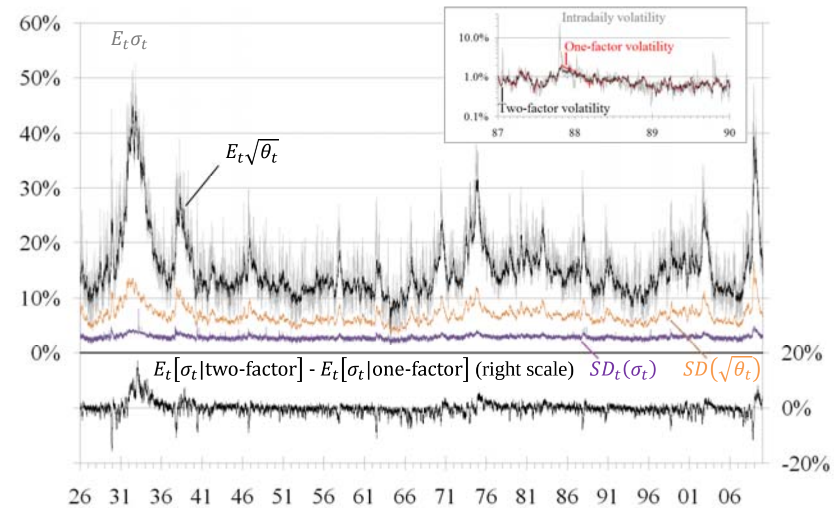
SVJ0: i.i.d. jumps in y_{t+1}

SVJ1: y_{t+1} jumps with intensity $\lambda_1 \sigma_t^2$

Filtration is different if σ_{t+1} jumps synchronously with y_{t+1}.

Bates (JFE, 2012) two-factor model of variance and crash risk evolution ($\lambda_t = \lambda_1 \sigma_t^2$)

σ_t^2 mean-reverts quickly towards θ_t (half-life = 1 week), which mean-reverts towards $\bar{\theta}$ (half-life = 1 year).



-Filtration can use other potentially more informative data sources. E.g., high-low ranges, intradaily realized volatility.
-I use it to assess whether volatilities inferred from index option prices (VIX) are accurate.